NPCR Education and Training Series (NETS)

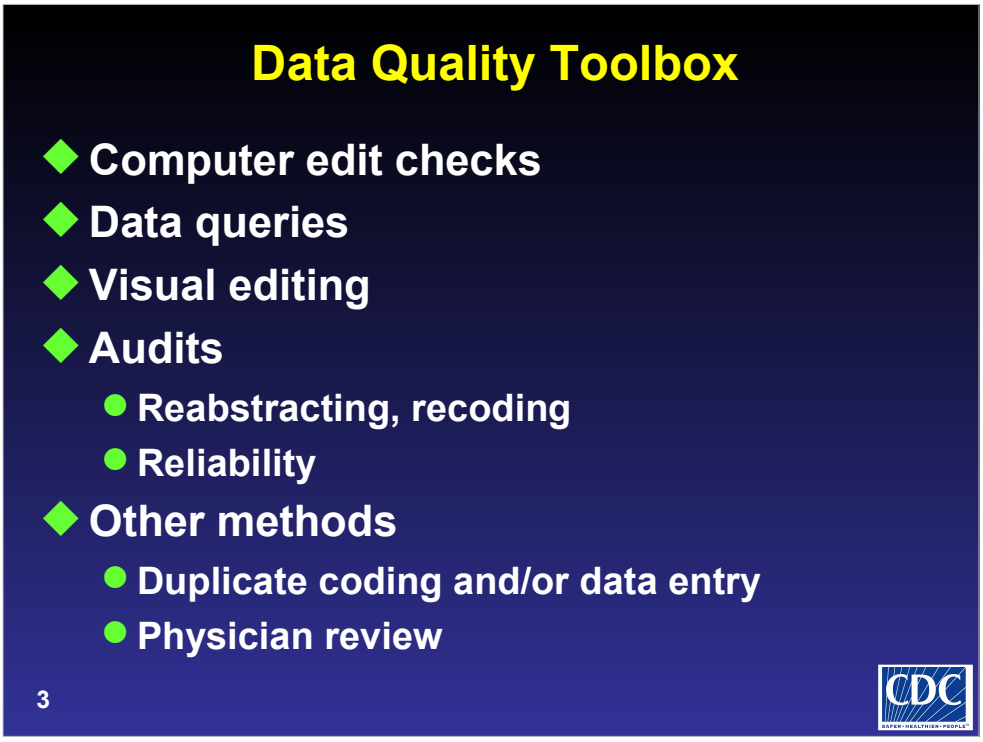
Module 3: Quality Control for Central Registries

Part 1-Section C: Registry Data Quality Control Tools

Prepared by Scientific Applications International Corporation (SAIC)
CDC Contract Number 200-2002-00576-0004

Centers for Disease Control and Prevention
National Program of Cancer Registries
Atlanta, Georgia
www.cdc.gov/cancer/npcr

Registry Data Quality Control Tools



Quality control staff in a central registry can use a wide variety of tools to monitor the completeness, accuracy, and timeliness of cancer data. A number of these tools were mentioned in the previous session. Each has value for specific types of registry data. This particular drawer of the proverbial data quality toolbox works with individual data items in individual records. Later, we'll talk about quality control measures for aggregate data and monitoring central registry data patterns.

Computer Edit Checks

- Range
- Inter-field
- Inter-record
- Inter-database

- May not be able to check for
  - Subtle inconsistencies
  - Incomplete records
  - Illogical information

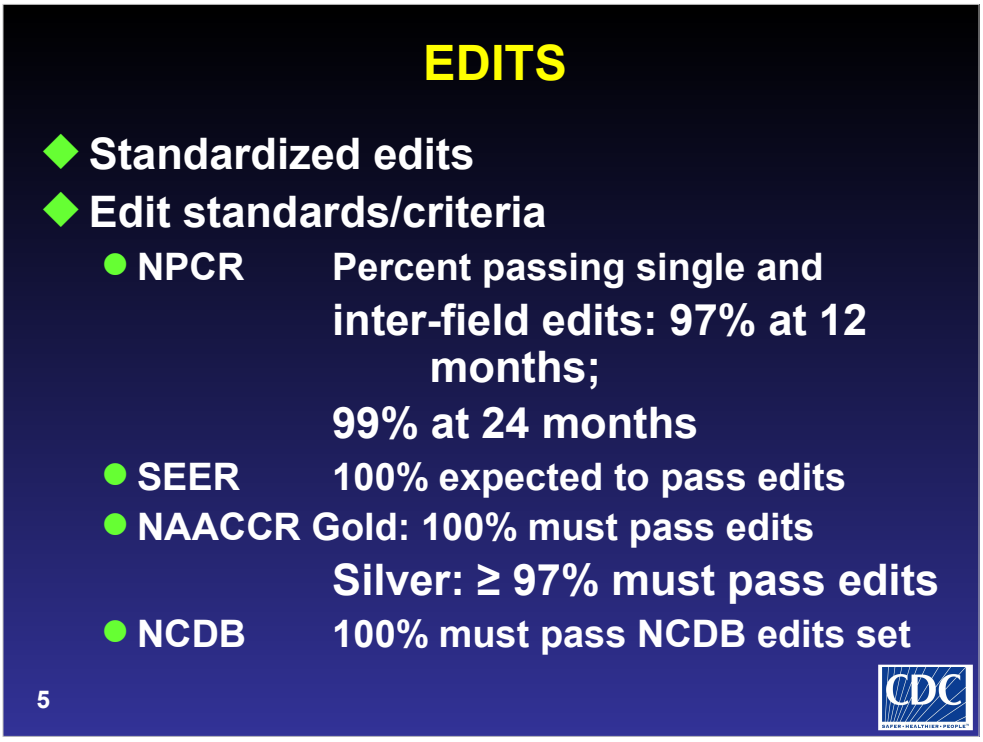
4

Computer edit checks are by far the most cost-effective way to monitor data quality. A well designed set of computer edits can accomplish in seconds what it would take a Quality Control (QC) staff member hours or even days to complete. By using computer edit checks, the central registry can clear many cases through baseline edits, freeing up the staff to take on the more complex, subtle, or sophisticated aspects of quality control.

Computer edit checks are a type of acceptance sampling, the process of inspecting, accepting, or rejecting a batch of data. In industry, acceptance sampling is performed on—as the name indicates—a sample of the product. In cancer registries, computerized edit checks should be performed on 100% of the registry's cases.

The four different types of computer edit checks have different purposes, and we'll discuss each in a moment.

Computerized edit checks are as smart as the person who wrote them, but they cannot find every incongruity in data. (Items may pass edits if coded to unknown, if there are inconsistencies or illogical information that the computer has not been programmed to detect, or information is missing.)

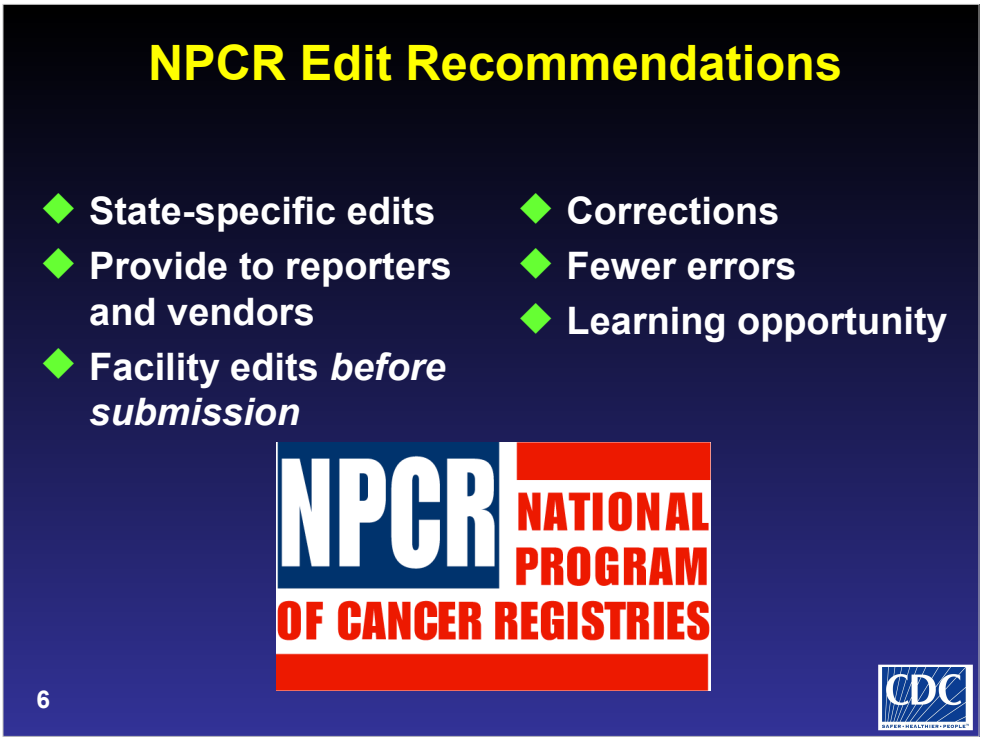


EDITS stands for Exchangeable-edits, Data-dictionary and Information Translation Standard.

CDC's NPCR is responsible for the development and maintenance of the EDITS software. NPCR also funds and supports the development and maintenance of the EDITS Metafiles.

The EDITS Software Project began in 1990 when registry software developers identified a need for standardized executable computer edits on registry data fields. EDITS is a set of software tools that can be used to improve data quality and standardize the way data items are checked for validity. The EDITS tools are available in a variety of applications both interactive and batch process. EDITS provides software to support three types of data activities, defining standards for data quality, standardizing data collection processes, and analyzing data quality.

Each national organization or agency has created its own set of computerized edits to be run on data submissions and each one has its own standard for passing edits.

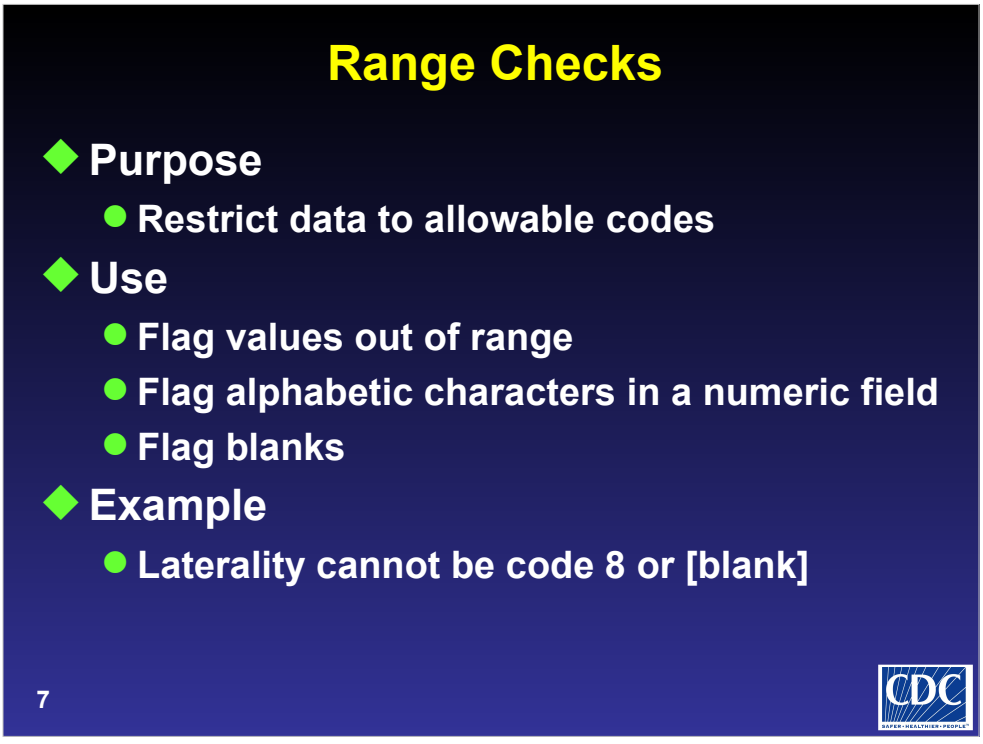


NPCR **strongly recommends** that each state develop their own state-specific edits; provide the edits to reporters and registry software vendors; and require their use. Each facility in your state should be running edits that meets your state's criteria *before submitting their data to the central registry.*

Editing data before submissions is beneficial to both the facilities and the central cancer registry.

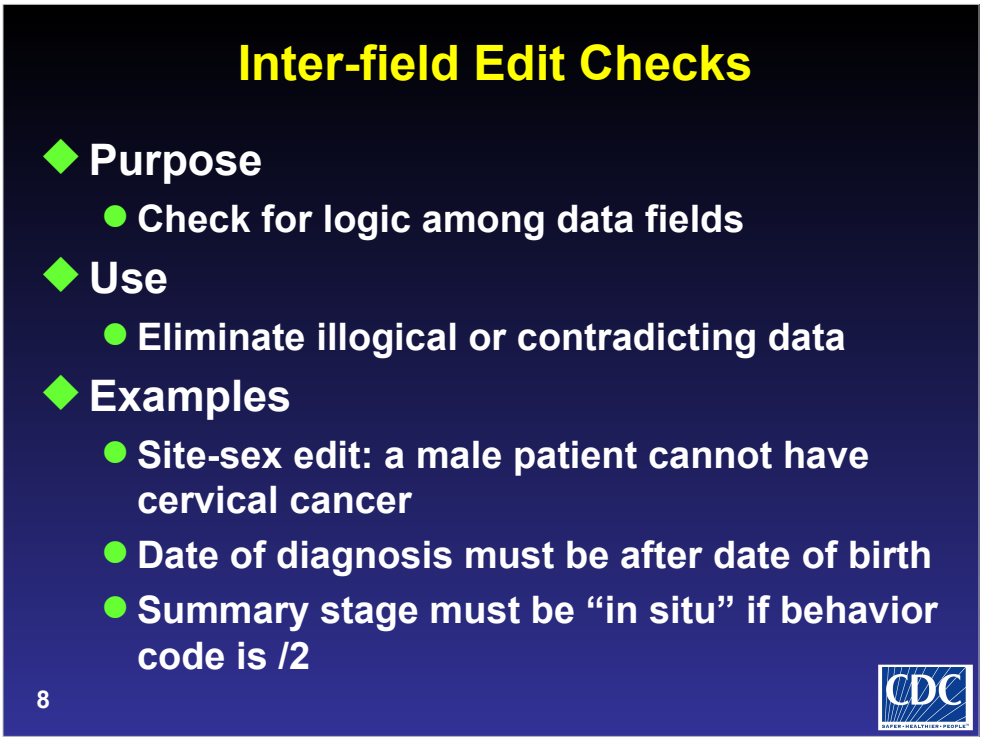
- For the facility, corrections can be made in a timely manner with immediate feedback to the abstractor(s).
- It means fewer corrections at the central registry saving time (perhaps saving time for facilities in that they would not be receiving as many calls from the central registry).
- Edits are always a learning opportunity and editing data before submission provides that learning opportunity in a very timely manner.

Now we will discuss the various types of edits.



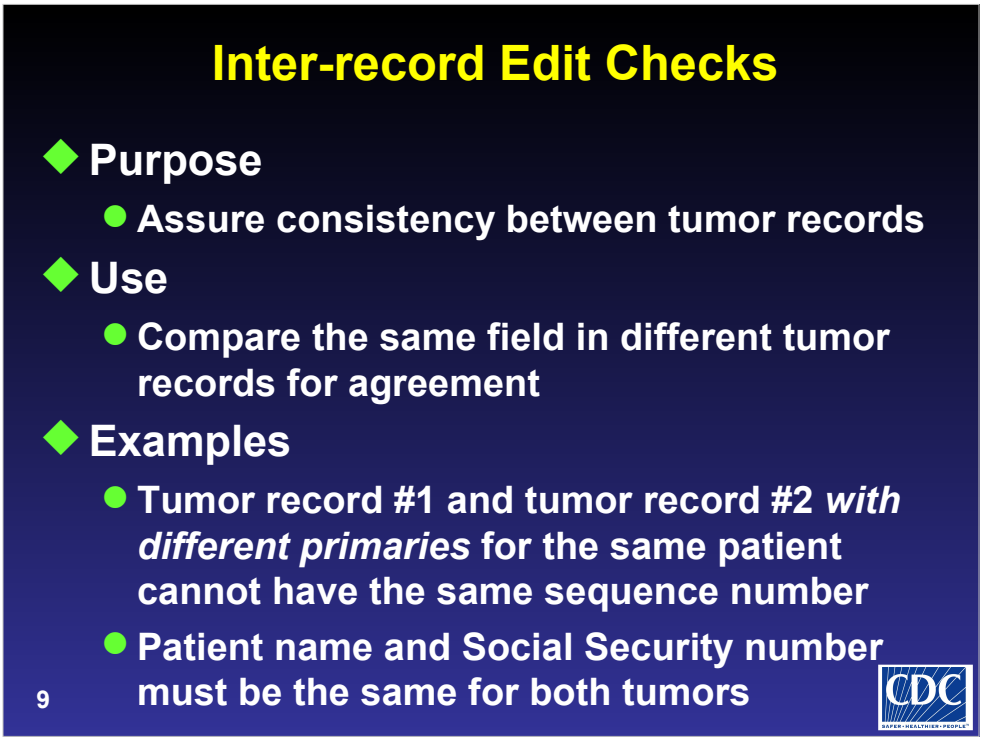
Range checks are single field edits. The purpose is to identify any field having a code other than those defined as acceptable for that field. Other names for range check are allowable code edit check and data format edit. For example, the laterality field is a single digit, so it has a capacity of 10 defined codes (0–9). However, the following six codes are the only acceptable codes for this field: 0, 1, 2, 3, 4, and 9. A range check for the laterality field would check for values of 5, 6, 7, and 8 as well as a letter, blank, or empty field. If any of these codes are found, the field fails the edit.

Range checks can be fairly detailed, depending on the data field. Values that can be used in a date field is another example. A range check may allow numbers 01–12 in month, 01–31 in day, a four digit year, and “smart” enough to know there are only 28 days in month 02 except in leap years. The range check may also be constructed to allow values representing unknown, such as 99 as the month or day, or 9999 for year. For certain fields, other non-date values may be allowed in the field, such as 00/00/0000 for none or never, and 88/88/8888 for not applicable. However, a true date field should never contain alphabetic characters. Range checks usually cannot be overridden by setting a review flag in the software; they must be fixed.



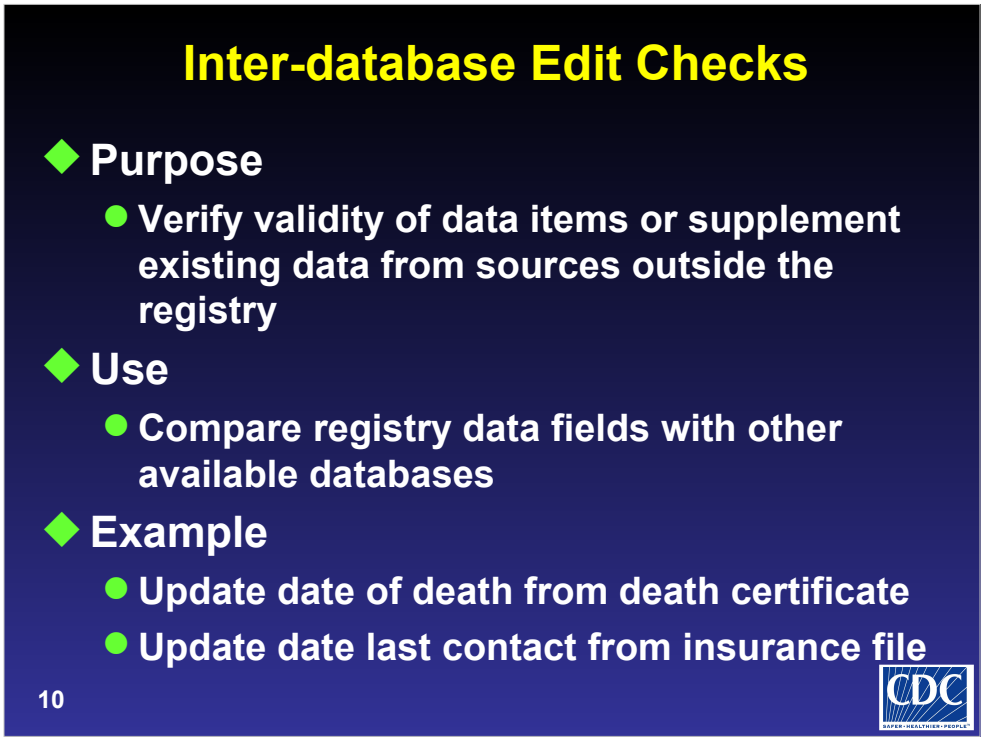
Inter-field edit checks compare the values in two or more data fields for logic. Inter-field edits are performed on a single patient-tumor record. Most standard edits do not look at the content of text fields as part of the logic check.

There are literally dozens of data combinations that have to make sense on an abstract. For example, treatment dates should logically be after the date of diagnosis. Certain types of cancer are not usually found in young patients (prostate cancer in a patient under age 40). Patients who have surgical resection (Surgery of Primary Site codes in the 20–90 range) should be histologically confirmed. Of course there are rare exceptions to inter-field edit checks like these. The FORDS manual and the NPCR data fields include a series of override flags to indicate that some combination of data field values may not seem logical but have been checked out and are true.



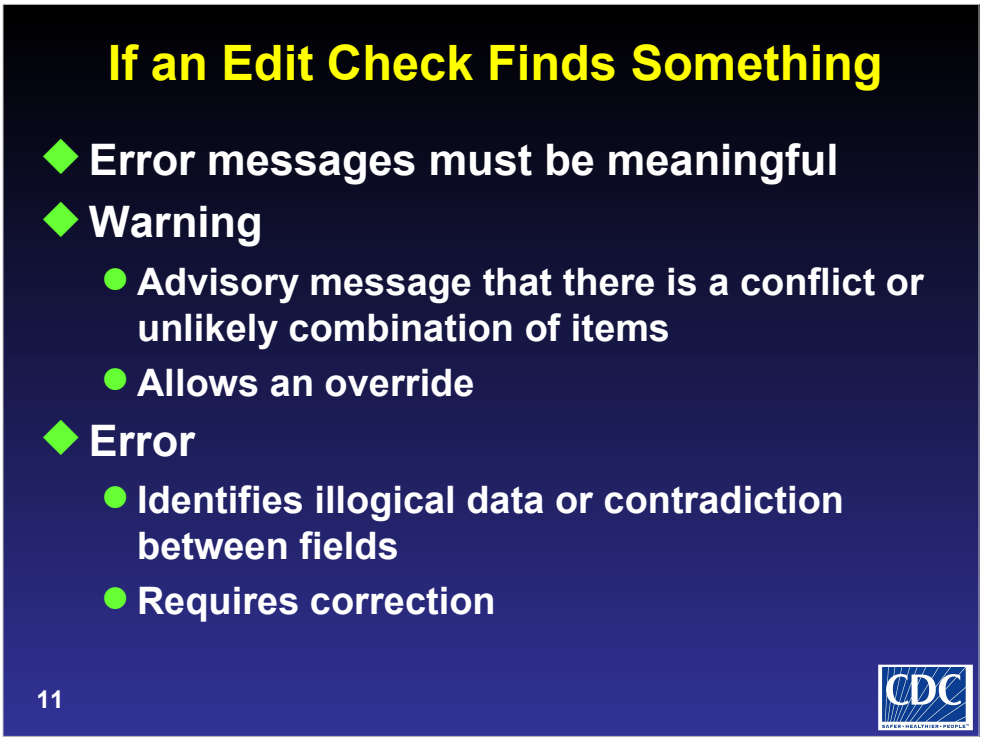
Inter-record edit checks compare data fields between tumor records for the same patient. These edit checks can run on tumor records from different facilities for a single primary in the same patient as part of the case consolidation process, or between tumor records of different primaries on the same patient in the registry database.

The classic example of an inter-record edit check is that a patient cannot be reported as alive at last contact on one tumor record and deceased on another tumor record. Such a conflict must be resolved. Similarly, there should be no discrepancies for patient Social Security number, date of birth, sex, race, or other patient identifiers.



Inter-database edit checks have several purposes. They are primarily used during data linkage with files outside the registry. An inter-database edit check can flag discrepant information and can even decide which information is better. For example, there may be a death certificate for a patient in the registry who is currently reported as alive. If all the patient identifiers match and the reported date of death is after the date of last contact on the registry file, the edit check can flag the case for updating. If the date of death is before the reported date of last contact, the edit check can flag the case for further investigation. Similarly, a death certificate may provide a specific race category for a patient whose registry record lists race as unknown.

A great deal of judgment is required of the quality control person reviewing inter-database edit checks. It may not be possible to know which of two or more discrepant facts about the patient is “true” without further investigation.

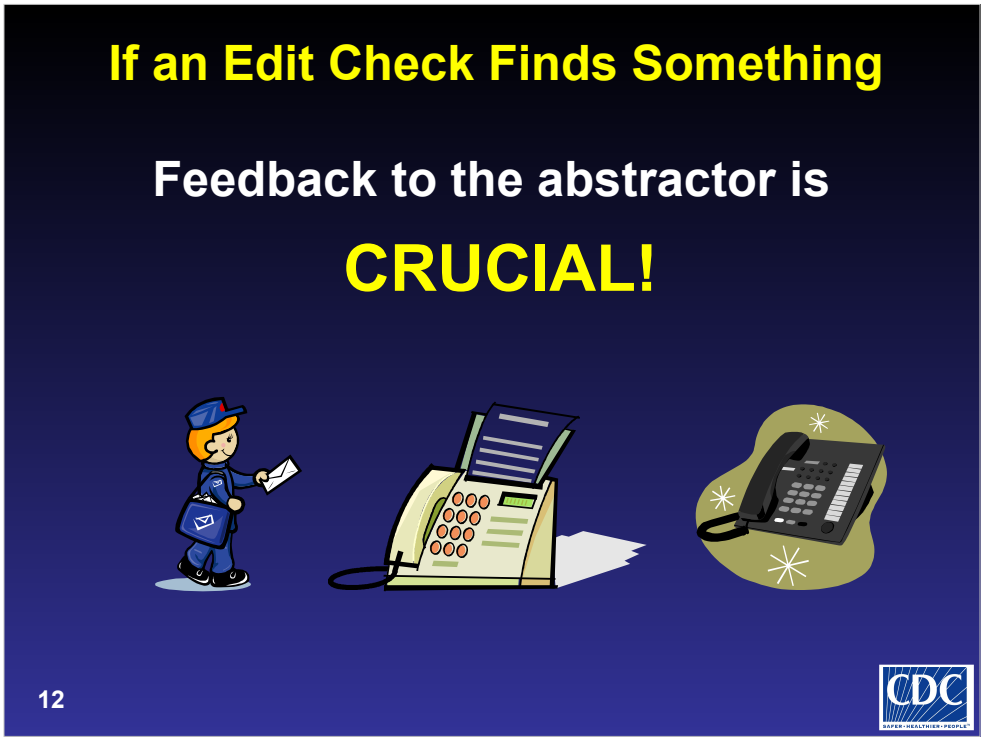


When an edit check finds a discrepancy or invalid data in a record, action must be taken. The edit check may trigger a warning or advisory message that the combination of data items is unlikely, or it may trigger a message that the conflict must be resolved.

Edit check warnings signal further investigation of the inconsistencies. A 36-year-old patient actually may have prostate cancer. If the information has been investigated and found to be unusual but true, an edit check warning allows the abstractor or quality control person to set an override flag.

An error message indicates there is an impossible combination of data fields or that there is invalid data in one of the fields. Some registry software programs will identify the fields or even jump to the field(s) requiring review and updating. Error messages require correction before the case passes.

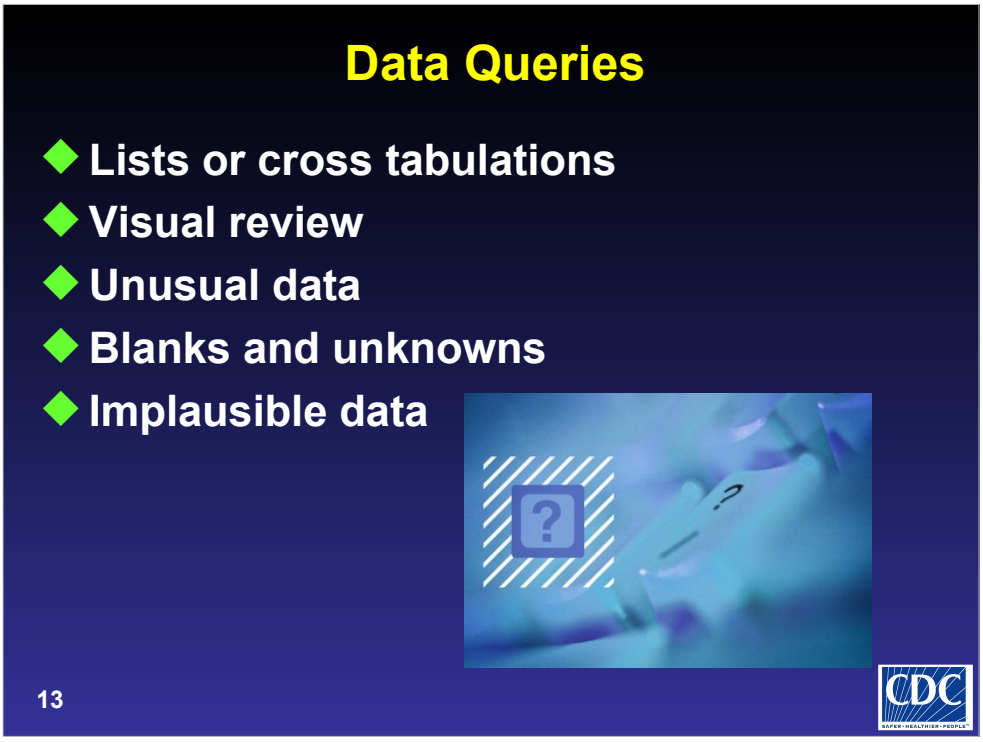
Your software may not differentiate between “warnings” and “errors” and may label both simply “errors.” In that case, some errors will have overrides and others will not.



If the edit checks are run at the central registry, feedback to data collectors on the types and patterns of errors identified during quality control activities is CRUCIAL. The feedback is part of the data collector's continuing education and can prevent or at least reduce the occurrence of similar edit check issues in the future.

The data collector/reporter will never know what needs to be corrected unless the central registry provides them with information about both the error and the correction. Yes, this does take some time and resources on the part of the central registry, but it is the "give" part of the give-and-take of data collection. The reporter may choose to correct the facility database to make it compatible with the central registry, but the main purpose is to educate the reporter.

Feedback to the abstractor provides the central registry with an opportunity to interact with the data collectors and is an acknowledgement that their data is important to the central registry.



Data queries are software programs that allow the user to look for specific features of cases in the database. Data queries can generate lists of cases by case identifier and data field as well as cross-tabulations of data fields. Usually the list or cross-tab requires visual review by the quality control staff. Data queries allow for a closer look at items for which there may be no edit check, or for which human judgment is necessary. For example, it is very unusual for a tumor size to be more than 10 centimeters (code 100). A data query can search for cases where CS Tumor Size is greater than 100 and less than 990. A printout of cases can be generated for review. The reviewer can request a printout of cases with lists of primary site, case identifiers, tumor size, method of diagnosis, one or more text fields, and other data items that would be helpful in validating tumor size. In some cases, such as an ovarian carcinoma, the large tumor size may be correct. In other cases, numbers may be transposed and the data needs to be corrected.

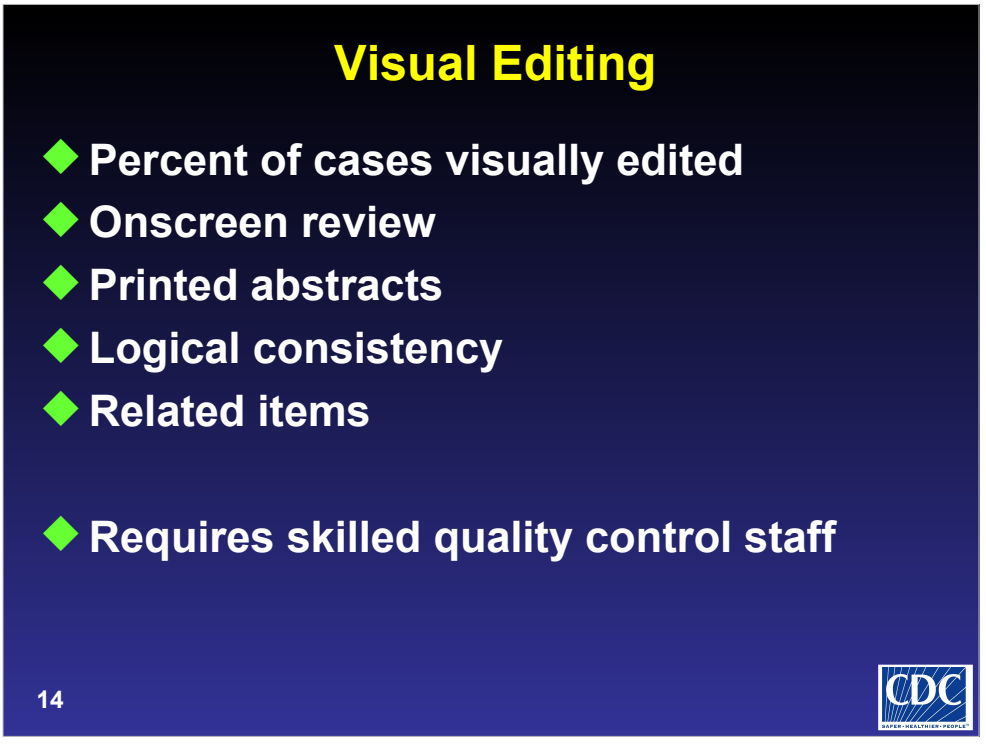
Visual review of a list can also show missing data and implausible data not found on computer edit checks. Some fields allow blanks, such as AJCC Stage Group. It's relatively easy to see a "hole" in a list where a data item is blank. The visual review will identify those cases for further investigation. A sorted list can also indicate the percentage of cases that pass computerized edits but are actually coded as unknown. Monitoring the frequency of unknown values is an important part of quality operations.

Data queries can be very creative. A query may be triggered by a news event or a question from a researcher. The point is that data queries offer the flexibility of searching the database for combinations of data fields that may pass edits, but may not make sense.

If inconsistencies are found that cannot be verified with text information from the abstract, the items should be followed back to the submitting facility for review and comment. If corrections are made in the central registry database, the submitting facility should be notified of the change so that its database can be corrected as well. [*Stress this point as part of central registry communication with its reporters!*]

NAACCR has developed a series of quality assessments that can be performed on any central registry database (http://www.naaccr.org/filesystem/pdf/CINA%20File%20Review%202006.pdf). These assessments look for excess use of override flags and inconsistent coding.

Data queries can be structured and saved, or ad hoc. In a central registry quality assurance program, there should be scheduling AND documentation of periodic review of the database using data queries.



Visual editing is the process of manually reviewing data fields on an abstract for consistency and logic. Visual editing can compare coded data fields to supporting text documentation, something that computer edit checks cannot do well. Visual editing validates the quality of abstracted data by comparing text and codes.

Visual editing standards vary from state to state. Some registries perform 100% visual review of abstracts, others review a smaller percentage of cases, based on budget, availability of quality control staff, and other factors. The percentage of abstracts reviewed might even vary for new registrars versus experienced registrars.

Some central registry software systems have built-in visual review mechanisms where data fields can be reviewed side-by-side for consistency. Others prefer printing out the abstract and reviewing the paper document.

Visual editing looks for logical consistency among data fields and verifies that consistency by reviewing the supporting documentation. The importance of the supporting documentation cannot be emphasized enough, as it is what permits the record to be edited without going back to the source documents for confirmation. Examples of related items will be shown in a moment.

It almost goes without saying that visual editing requires highly skilled quality control staff. Most often, visual editors are experienced hospital registrars who have joined the central registry. A visual editor must know data field codes, coding rules, the relationships between data items, site-specific cancer disease processes, and expected types of treatment by stage. The work takes extreme concentration and a large dose of common sense, and the skills take years to develop.

Visual Editing

- Why bother?
  - To document problem areas
  - To quantify reliability of data
  - To make the abstractor think about interrelationships of data fields

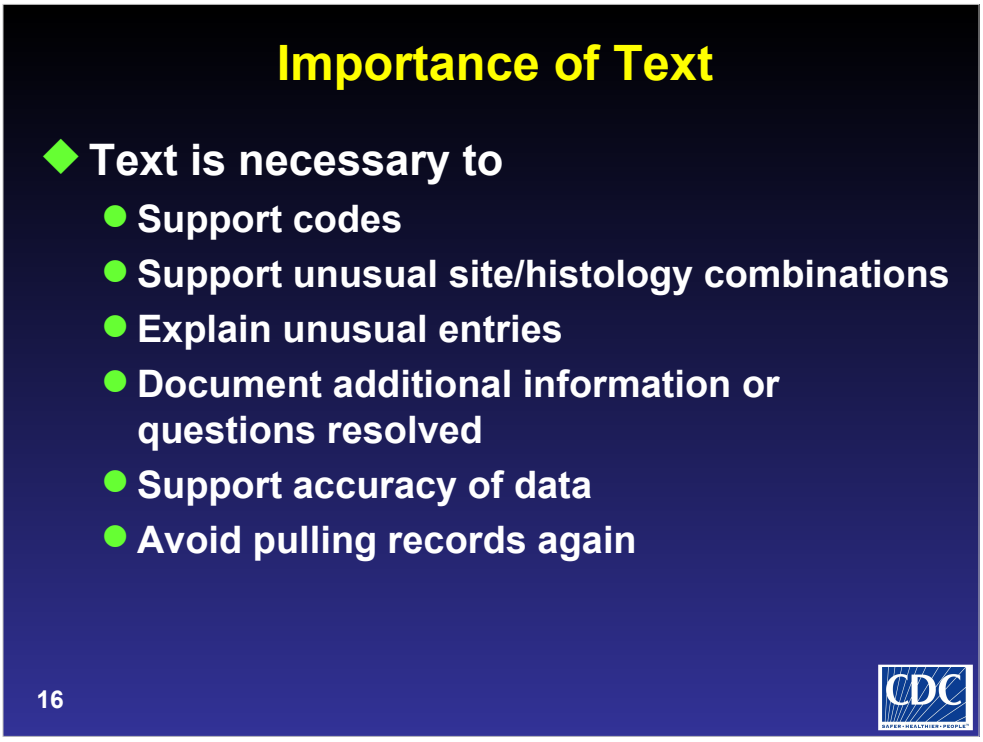
15

We know that there are specific data items that are critical for central registry use. Why can't the central registry just take the abstractor's word that the data are accurate?

The main reason to edit the data is to quantify the accuracy for the sake of data users (researchers, health planners, and everyone else).

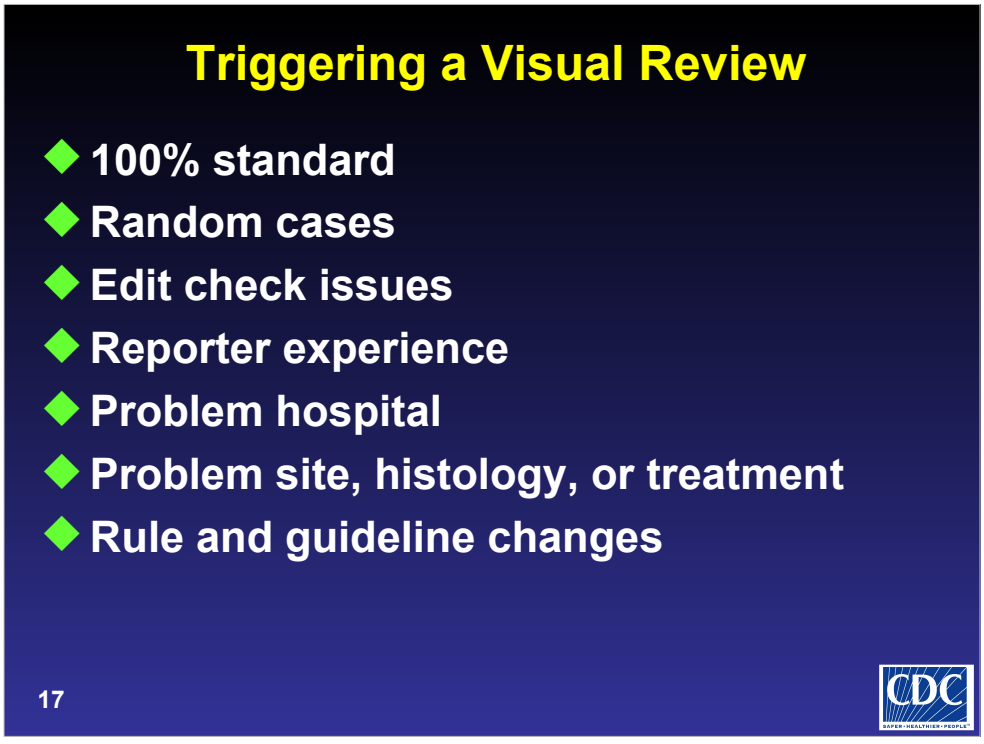
In the process, problem areas can be identified and the central registry can address the issues through guidelines, rules, and various types of training.

It never hurts to make sure that the data going into the system is as good as it possibly can be.



Visual editing is a way to cross-validate data because it is entered twice—once as code and once as text. The job of the visual editor is to review the codes and compare the meaning of the code to the text or narrative that supports it.

Many central registries do not emphasize the importance of abstractors submitting supporting text, but then they are at a disadvantage when it comes to quality-controlling the codes. Unusual combinations of site and histology or other fields look very odd to a researcher if there is no explanation or justification of the apparent discrepancy. In addition, text can document a subcategory of a code or a more specific fact that is represented by a non-specific code. To say again something we have already mentioned, text will validate or prove the choice of the code, making it more accurate and reliable. And not inconsequentially, proactive documentation will save retrieving the medical record again if a code is questioned by the central registry or a researcher.



Various factors may trigger a visual review. If the registry's standard is to review 100% of abstracts, then every case will be visually edited. If the registry standard is lower, a number of factors should be considered.

The registry may adopt a policy of reviewing a statistically random sample of all cases, such as every fifth case (e.g., 20% of cases) added to the central registry database. Or, any case that fails one or more computerized edit check may be flagged for visual review. Alternatively, the registry may target certain types of cases to be reviewed. For example, the registry may review 100% of cases submitted by novice abstractors (those with less than one year of experience) and review only 10% of cases submitted by experienced abstractors who have a track record of submitting high quality data. Often the policy is to review all cases submitted by novice abstractors until several sequential batches are error free. In the case of experienced abstractors, periodic spot-checking or review of random cases could identify an increase in error rates that may trigger a more complete review. Here again, acceptance sampling and statistical process controls can be used to set the thresholds for further review.

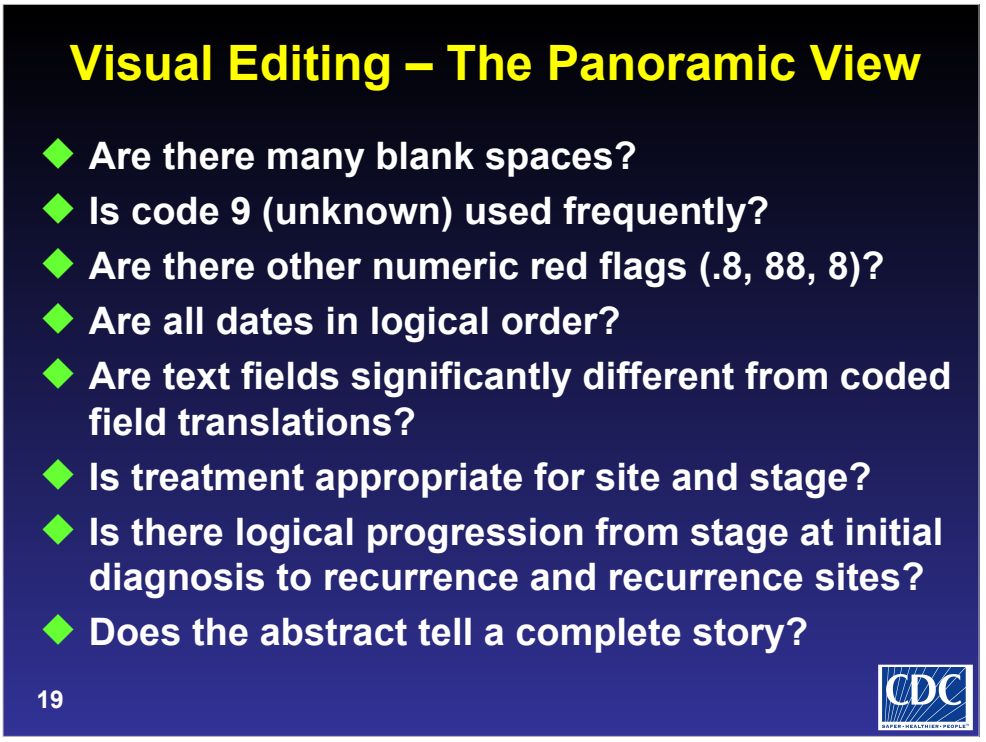
The registry may choose to review a higher percentage of cases from a hospital with staffing problems or a hospital that hires unsupervised contract abstractors. The registry may target problem cases based on previous audits, such as lung cancers or lymphomas, or patients who would be expected to have a particular type of treatment. Cases abstracted in the few months after a major rule change is implemented can also be the subject of visual review.

How does the process of visual review work? Let's take a look.

The next few slides and the notes that go with each one include very extensive lists of items for review during the visual editing process.

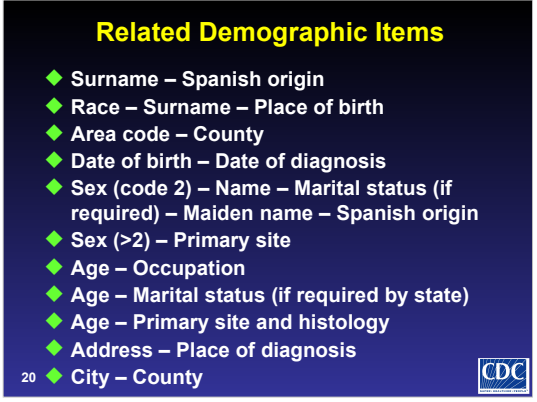
Please note that these slides and notes could be very useful in developing your procedures for visual editing and could easily be formatted into a guide or checklist for visual editors. Of course the value of established guidelines and procedures is consistency among the visual editing staff.

These slides and the notes can also be used to create part of your quality control plan.



The following considerations are based on visual review of a paper abstract. The same concepts would apply if visual review is conducted on a computer screen, but the number of fields will be more limited at each point due to the constraints of the computer screen.

Before you begin looking at individual fields, take a look at the panoramic view. This is the initial impression of the case. Review these questions as you look at the case. Is there anything that raises a red flag? Codes 8 and 9 will pass edits, but are they a signal of incomplete data? Does the text say sarcoma and a transposed histology code is translated as carcinoma? Remember that this first review is general, the specifics will arrive in a minute.



This is a partial list of related items that can be compared during a visual editing session. It just takes a split second to look at these, assuming that there is data in the fields.

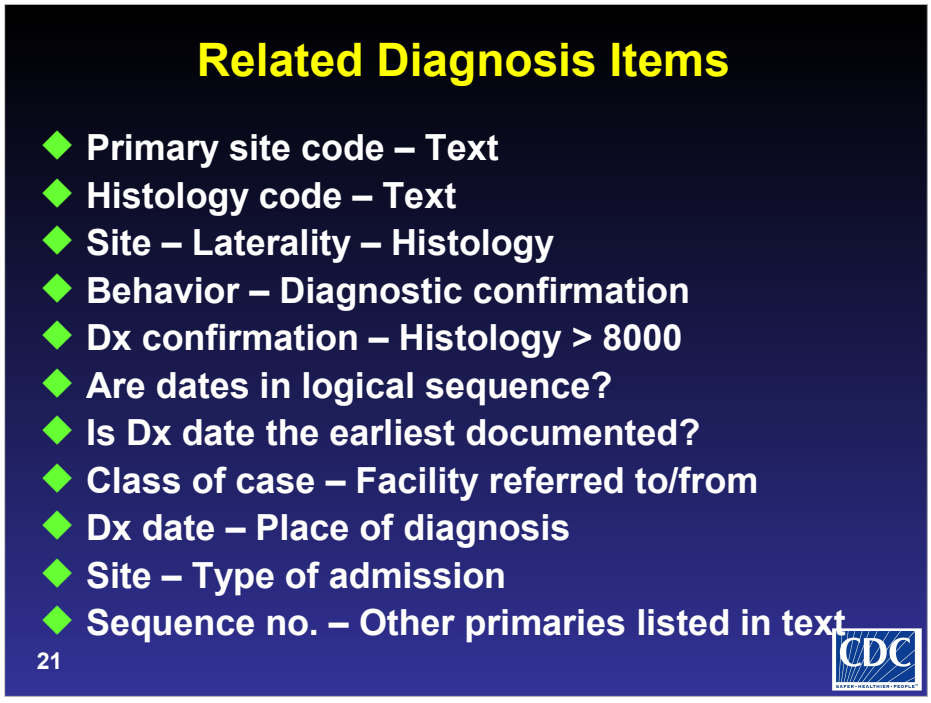
The demographic fields include the basic personal identification of the patient and critical information for incidence reporting. Individual items are important too, not just pairs to compare.

For example:

- Are standardized abbreviations used in the address?
- Does the Social Security number show all 9 digits?
- Does it start with an 8 or a 9?
- Does the patient's name correspond to the coded ethnicity? Name lists are available for Hispanic and Asian surnames.
- Does race correspond to name and birthplace?
- Does race correspond to primary site? For example, darker pigmented populations rarely develop skin melanomas, and Chinese have higher rates of nasopharyngeal and primary liver cancer than the general population.
- Does the area code of the patient's phone number correspond to the county of residence?
- Think about first names:
  - — Kaposi's sarcoma in females is infrequent
  - — Breast cancer in males occurs in about 1 in 100 breast cancers
- Check that the date of birth and the date of diagnosis are not transposed or the same. If the same, make sure that the primary site and histology are appropriate to an infant and add a remark to verify patient is a newborn.
- If the patient is female and married, does the married name or maiden name correspond to the coded ethnicity?
- If sex is coded 3 or 4 (big red flag).

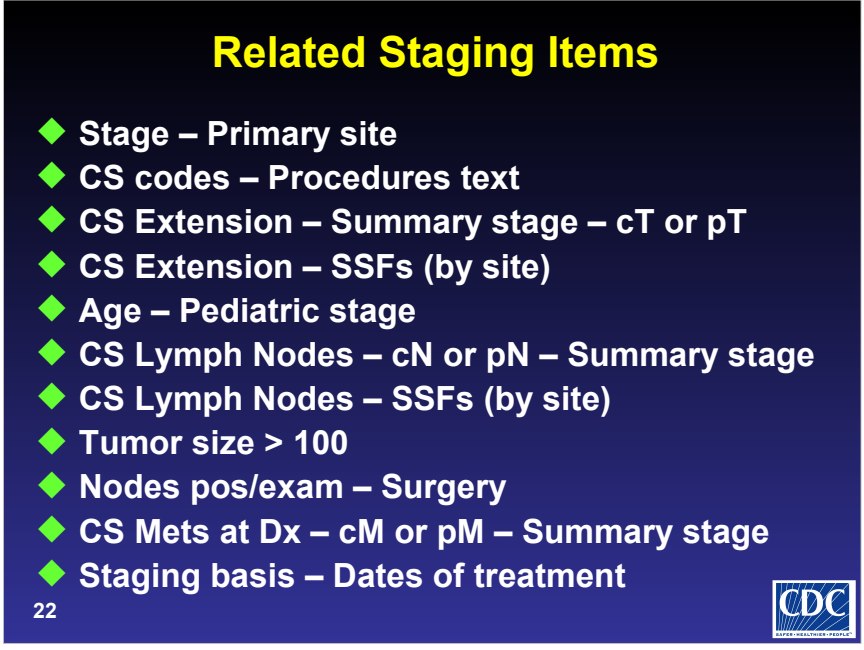
Keep in mind that:

- Hermaphrodites usually develop genital cancers such as ovary or testis.
- Transsexuals are at risk for Kaposi's sarcoma and lymphomas.
- Monitor the age carefully. Any age 000 or 100 should be a red flag to be checked and verified. Any age < 30 or above 90 should be checked.
- A child (15 or under) would not normally have an occupation other than “child” or “student.” Also, check for a parent listed somewhere on the abstract.
- A child (15–18) would not normally be married, widowed, divorced, or separated. If married, document in remarks.
- A child would not normally have an epithelial cancer. Children develop leukemia, lymphoma, astrocytoma, neuroblastoma, sarcoma, hepatoblastoma, Wilms' tumor, or retinoblastoma.
- In contrast, nearly two-thirds of cancers cases occur in patients aged 65 years and older. Adults rarely develop retinoblastomas, Wilms' tumor, or bone tumors.
- Super-annuated (elderly) patients usually develop prostate, lung, breast, or colorectal cancer.
- Is there a difference between where the patient resides and where she or he was diagnosed or treated? This might be a signal that the patient has more information at a facility closer to home, or that the patient is a nonresident of the area.
- Is the city in the correct county?
- And by the way, is the ZIP Code numerically correct and within the range for your area?



Diagnosis fields include some administrative fields and the majority of supporting text fields.

- Does the translation of the code say nearly the same thing as the text? Does the text justify the codes?
- Does the histology code translation say the same thing as the pathology text? Do the primary site and histology make sense together?
- Primary site – is it documented in the text? Which text?
- Is the site code correct?
- Is the subsite specific or NOS? Is an NOS code validated?
- Is a .8 site code described in the text? What does the tumor overlap?
- Does the site and histology combination make sense?
- Some site-histology combinations intuitively go together, such as prostate with adenocarcinoma, head and neck with squamous cell carcinoma. Know too what site-histology combinations are unlikely, such as adenocarcinoma of the upper and middle esophagus, or carcinoma in the bone or spleen.
- Is this a paired site, and is the laterality documented in the text?
- Does behavior correspond to diagnostic confirmation? Is this an in situ case diagnosed clinically or by cytology? Or maybe a multiple myeloma diagnosed on electrophoresis.
- Is there a specific cell type coded if there is microscopic tissue confirmation?
- Does the differentiation code agree with the text, for example code 3 for moderately to poorly differentiated tumor?
- Are dates of diagnostic procedures before date of treatment? Are excision dates after needle biopsy dates?
- Is the diagnosis date the earliest date a recognized medical practitioner said the patient had cancer? Is that date between the admit and discharge dates?
- Date of diagnosis – is it clear how and where the patient was diagnosed on this date?
- Do all of the dates lead up to the tissue diagnosis and definitive treatment?
- Do the first four digits of the accession number reflect the date of admission of the first primary cancer for this patient at the facility?
- If the patient is not a class of case 1, are hospital referred from and hospital referred to filled in as appropriate? (Referred from for class 2 and referred to for class 0.)
- Does the diagnosis date correlate to the place of diagnosis?
- Is the type of admission appropriate to the primary site? For example, is the patient having brain surgery as an outpatient?
- If the sequence number is greater than 00, are any other primaries documented in the remarks space?
- What about number of reportable by agreement cases and/or benign and borderline tumors collected by the registry?
- And by the way, if this case is sequence 02, does the previous abstract say sequence 01, rather than sequence 00?



NPCR, SEER, COC, and NAACCR studies have shown that the staging fields are the most prone to problems, because coding these fields requires interpretation and sometimes the codes or the rules are rather subjective. However, these fields are vital for monitoring cancer trends and treatment. The elements of Collaborative Staging are for the most part facts, and they can be mapped to other stages. If the other staging fields are completed by physicians, this is another way of cross-checking the data.

- Does the stage correlate to the primary site?
- Do you have an in situ sarcoma?
- Do you have a regionalized leukemia?
- Do you have a localized unknown primary?
- How about a Stage IV melanoma?
- Are the CS codes supported by text?
- Are the CS codes correct for the time frame, according to the dates of various procedures?
- Does the T (clinical and/or pathologic) correspond to summary stage and is CS Extension in the range 00–80? It would be unusual to have a physician-staged T1 and an Extension code of 50, or a summary stage localized with Extension in the 70 80 range.
- If patient is a child, is a pediatric stage (rather than an AJCC stage) shown?
- Does the N (clinical and/or pathologic) correspond to summary stage regional to lymph nodes or regional nodes and extension, and is the CS lymph node code > 00 if N1?
- Is tumor size an important criterion for this site? If so, does it have a valid number? Is tumor size greater than 10 cm? And is size documented in text?
- Is clinical stage based on information before treatment was started?
- Is pathologic stage based on adequate surgical and pathological information?

Related Treatment Items

- Planned first course listed?
- Treatment – Primary site – Stage
- Treatment code – Procedure name
- Treatment – Facility referred from/to
- Surgery – Operative findings text
- Surgery – Pathology text
- Date 1st surg – Date most definitive surg
- Date most definitive surg – Date surg discharge
- Surg prim site – Margins
- Surg prim site – Scope reg LN
- Surg prim site – Reason no surg

23

•The treatment fields are where you need to widen your focus. From the codes and supporting text, can you tell what the planned first course of treatment was, and what the reasons were if the patient did not receive the usual treatment for the site and stage of disease?

•Do the treatment codes, particularly the surgery codes, include the procedure named in the text? Compare treatment fields and treatment this hospital fields. If they don't match, is a referral hospital documented?

•Is the surgical procedure listed in both the surgery code area and the operative findings?

•Here again, are the dates of treatment in logical order?

•If there was a "big" surgery to the primary site, were lymph nodes removed as well?

•Was there a pathology specimen?

•Is Dx confirmation appropriate to the surgery and path information?

•Does the pathology report correlate to which organs were removed in the coded and named surgery?

Related Treatment Items

- Surgery – Radiation – RT/surgery sequence
- Date RT start – Date RT end
- Location of RT – Facility referred from/to
- RT treatment volume – Reason no RT
- RT treatment volume – Boost treatment volume
- Systemic tx – Primary site
- Systemic tx date – Chemo – Hormone – Immunotherapy
- Systemic tx date – Date most definitive surg – Systemic/surg sequence
- Hemat Transpl & Endocr Proced – Primary site
- RT treatment volume – Palliative care

24

•Is the radiation modality code supported by text?

•Are drugs listed in text for chemo, hormone, and/or immunotherapy, and are they coded correctly? For example, is Levamisole coded as chemo rather than immunotherapy as it should be? Is Rituxan coded as immunotherapy rather than chemo as it should be?

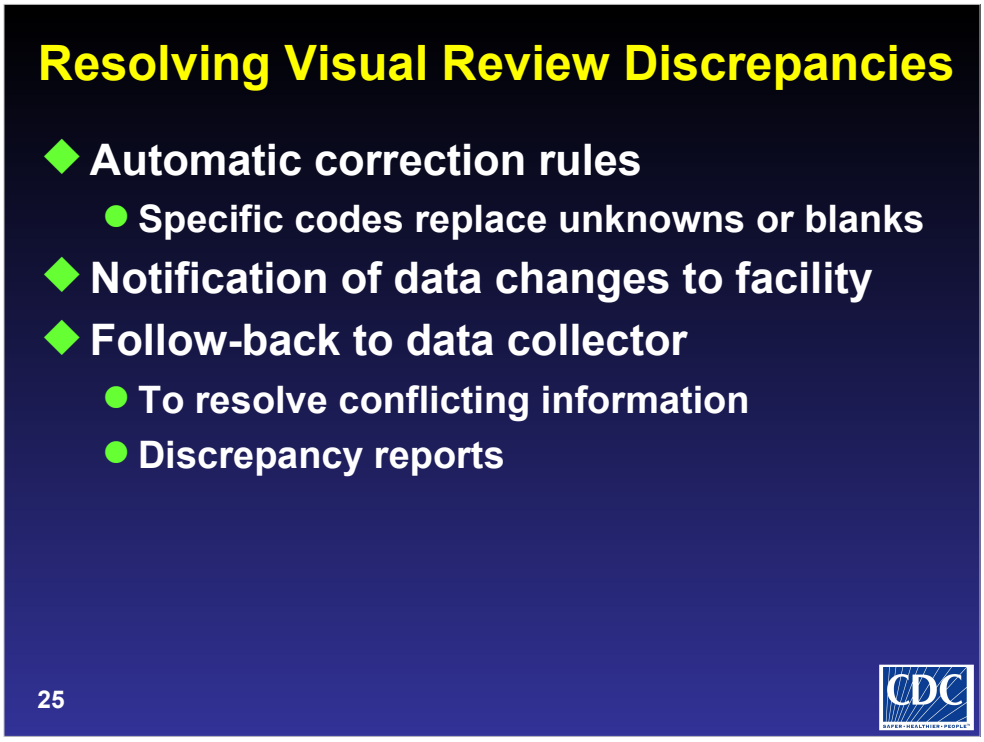
•If systemic therapy changed, was there disease progression documented (the change would be coded in subsequent therapy, which may not be on the central registry abstract) or was the change planned as part of first course (CHOP-ABVD or induction and maintenance therapy for leukemia)?

•Does the primary site correlate to systemic treatment? Is there an early stage cancer receiving chemotherapy without evidence of involved nodes?

•Do the dates of treatment support the codes for surgery/RT sequence and surgery/systemic therapy sequence?

•If any type of treatment was refused, is the reason documented in text?

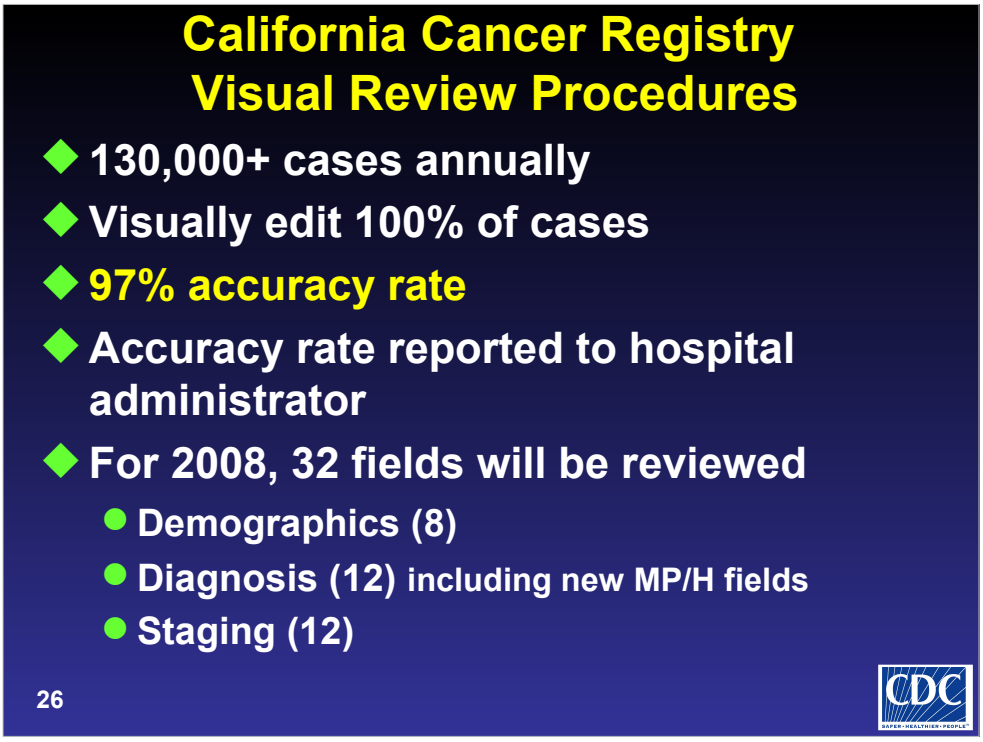
•Look at the radiation treatment fields—if they look like sites that would receive palliative treatment, is the palliative care field also coded?



When better information can be identified from the text, specific codes replace unknowns or blanks on the abstract. The central registry may also have other automatic correction rules.

Once again, it is crucial that any discrepancies identified in visual review be reported back to the facility as part of an ongoing quality improvement process for the reporter.

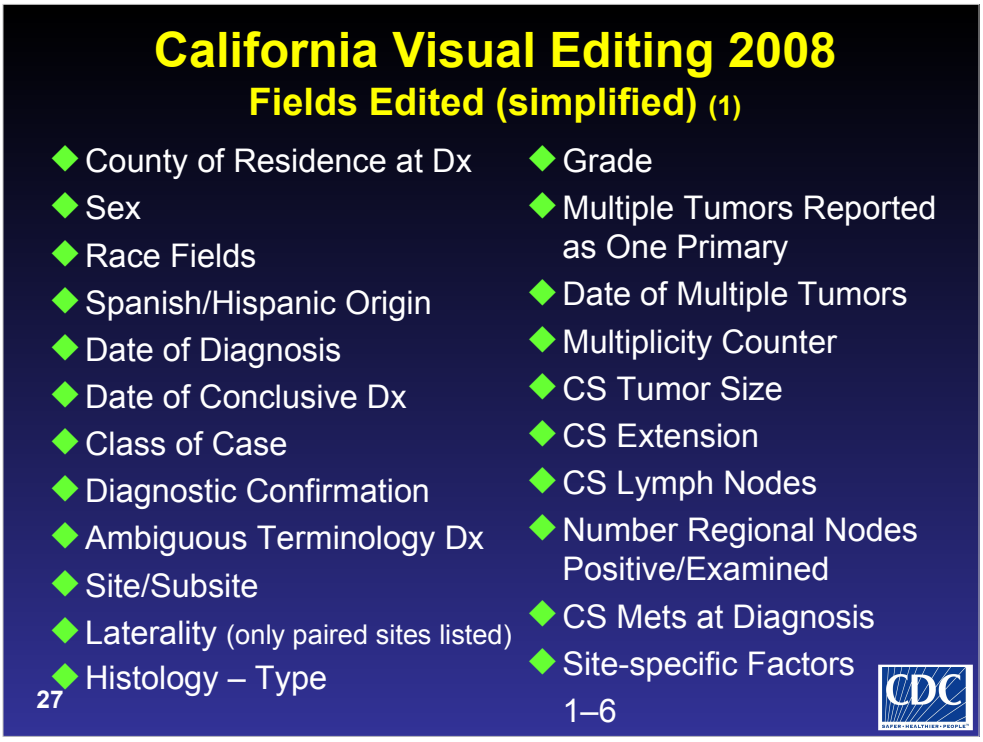
Queries back to the data collector should be used to resolve conflicting information as well as provide periodic discrepancy reports.



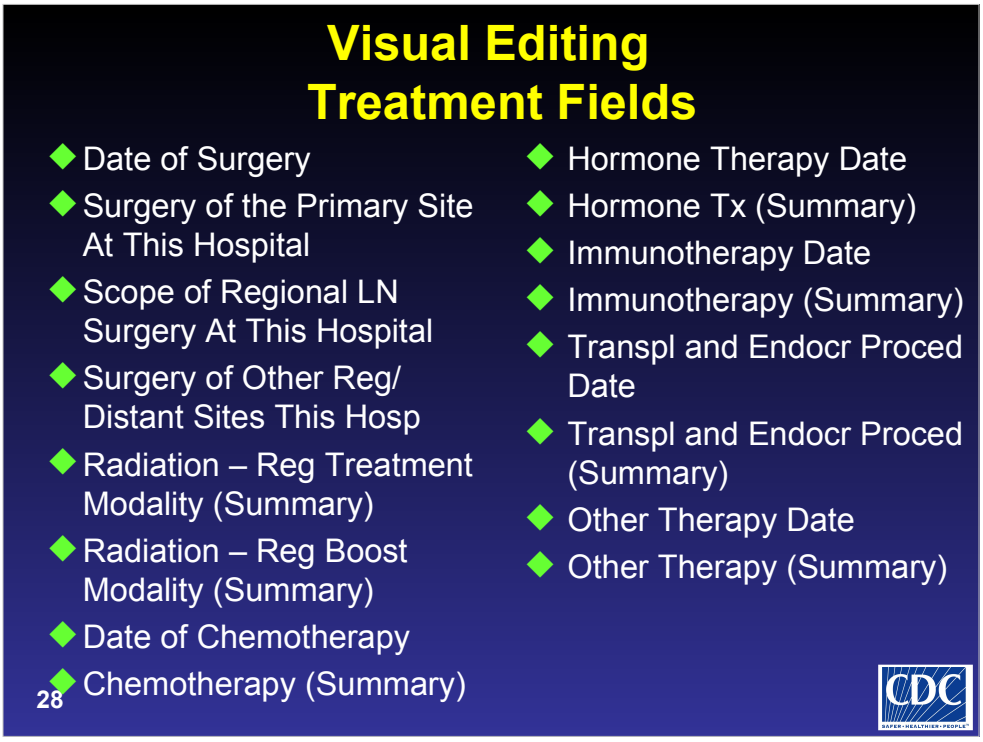
Visual editing is not a new concept. For example, **until recently** the California Cancer Registry has been visually editing 100% of its new cases, with more than 130,000 new incident cases added to the state database every year. That's a lot of visual editing.

In 2000, the California Cancer Registry and its 10 regions implemented the most stringent quality control standards in the world. The state established a 97% accuracy rate for 13 data items that would be visually edited by the regional registry. The accuracy rate would be reported to the facility's administrators on a periodic basis. Since 2000, the number of data fields as been expanded, and in 2008, 40% of the data will have 32 fields visually edited with that same expected 97% accuracy rate.

This standard of accuracy places a tremendous responsibility on the data collector and as much a burden on the visual editors, but the results have proven worth the effort. As of the first quarter, 2007, the AVERAGE accuracy rate for all hospitals is 98.6%!



These are the demographic, diagnostic, and staging fields that are visually reviewed in 2008 in California. Note that they include the fields added to document the multiple primaries and histology coding rules that were implemented in 2007. Note also that text fields are not included in this list but are required to support the coded fields listed.



The treatment fields on this slide can be added to the list of data items for visual review. Again, note that text fields are not included in this list but are required to support the coded fields listed.

If your central registry has never done visual editing, or if you are just starting this quality improvement technique, start with just a few visual edits, particularly those that affect incidence reporting. Or, pick 10 or 12 data items that are used most frequently in research in your state. As the visual editing process matures, you can expand the number of fields that come under scrutiny as staffing and resources permit.

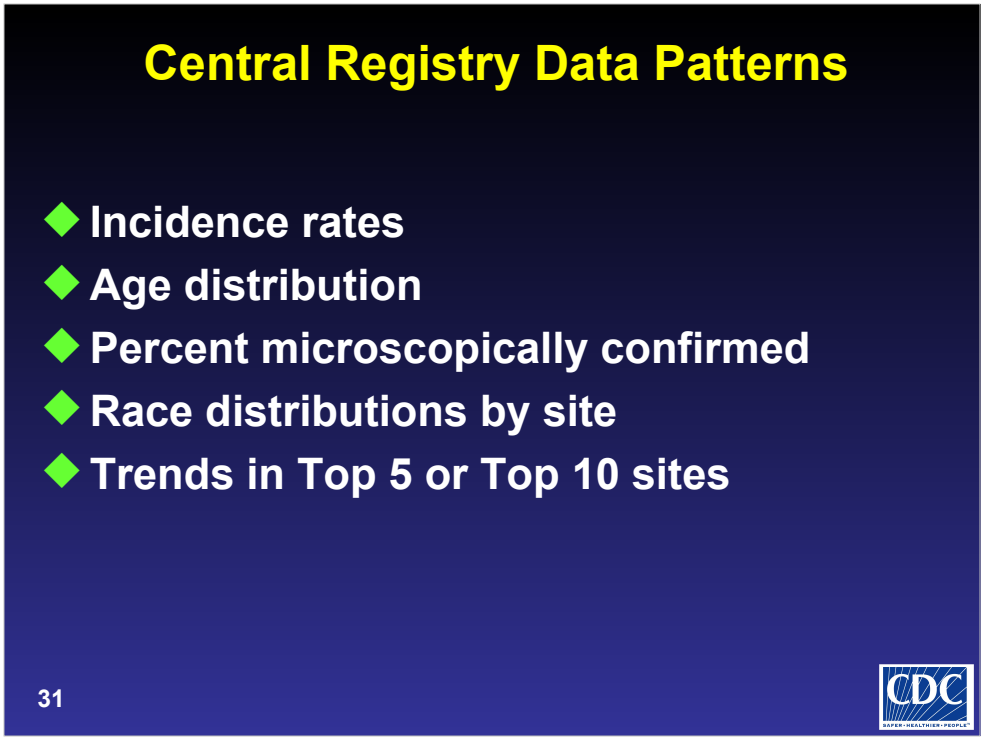
Other than computer edit checks and visual editing, there are other data quality control methods or tools that can be used.

Physician review of cases is a requirement in Commission on Cancer–approved facilities. At the central registry level it is more of a luxury and should be reserved for difficult, complex, or medically challenging cases. Any consultation with a physician advisor should be documented on the case at the central registry level.

Audits perform a vital role in quality control of registry data. Audits and designed studies can identify trends and gaps in training, but generally do not have an immediate direct effect on data quality. It is only over time—after the problems have been identified and training has been done—that audits can affect the data. Audits will be covered in later sessions.

In medicine, the Zebra Rule is "When you hear hoofbeats, think of horses, not zebras." Although physicians are trained to identify exotic and unusual diseases, they are also frequently reminded that most diseases are the common types.

In a cancer registry, that rule has two aspects. When you are abstracting, think horses. When you are editing, think zebras. In other words, when you are abstracting, think of common things that you would expect to be found. When you are editing, look for the unusual—things that don't make sense or stand out from the "usual."



In **central cancer registries** data patterns that should be analyzed to ascertain if they meet expected norms. Common sense and comparison to other central registries should be the bases for this type of quality review. The following examples show common sense observations and rules of thumb.

- Incidence rates and frequencies should be greater than mortality rates and frequencies. If mortality rates appear to be greater, this MUST be verified.
- Where the age distribution of the population at risk is similar to the national norm, childhood cancers should represent about 1% of the total number of cases. Cancers in persons over 80 years of age should account for approximately 15% of the total.
- In the U.S. the microscopically confirmed cases should represent approximately 93%. This will vary by state and even within states according to distance to a treating facility and other measures of access to medical care, such as socioeconomic criteria. However, every central registry should have a mix of pathologically and clinically diagnosed cases.

Certain race distributions should be observed by site. Examples identified over years of analysis include—

- Rates for cancer of the corpus uteri should be higher for whites than for blacks, and rates for cervix uteri should be higher for blacks than for whites.
- Significant numbers of melanomas occur only in the white population.
- Cancer of the prostate should be higher for blacks.
- Rates for Hispanics are generally lower than for whites except for cervix uteri, esophagus, stomach, and pancreas.
- Certain site-specific patterns should be observed.

Examples include—

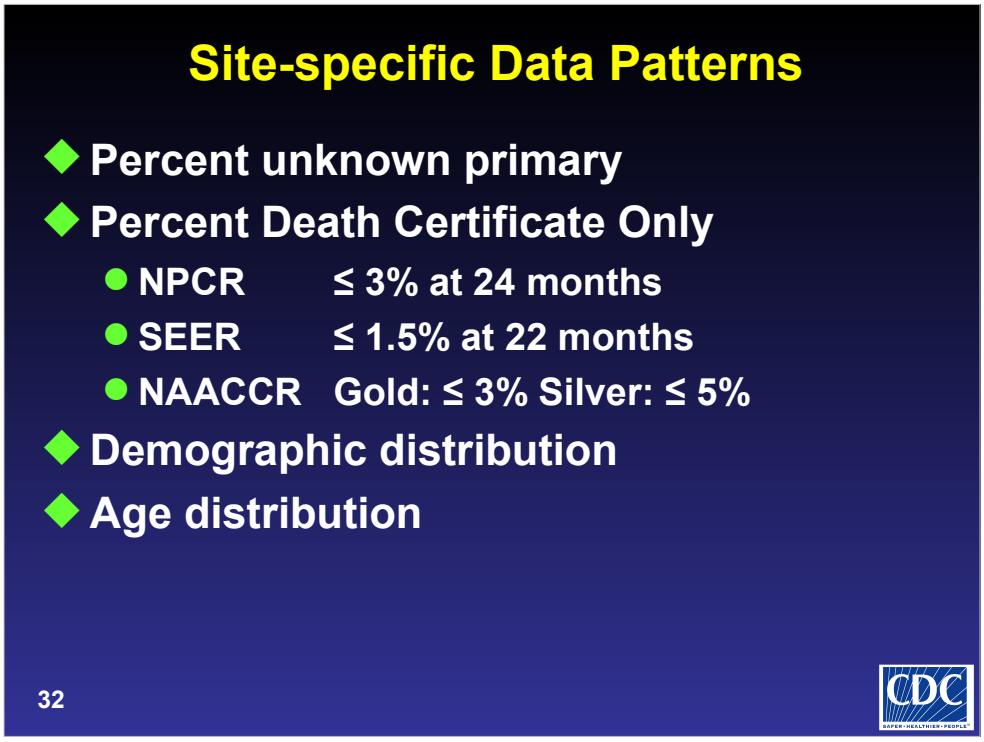
- Death certificate only cases for lung, liver, and pancreas.
- Unknown primary site should be about 5% of all cases.
- Male breast cancer usually accounts for about 1/2% to 1% of all breast cancer cases.

Trends in the top 5 or top 10 sites by sex will fluctuate slightly, but generally will include—

- Breast, lung, colorectal, corpus uteri, lymphoma, and/or ovary for women.
- Prostate, lung, colorectal, bladder, melanoma, and/or lymphoma for men.

Trends in the top 5 or top 10 sites by race or ethnicity will also fluctuate, but generally will include—

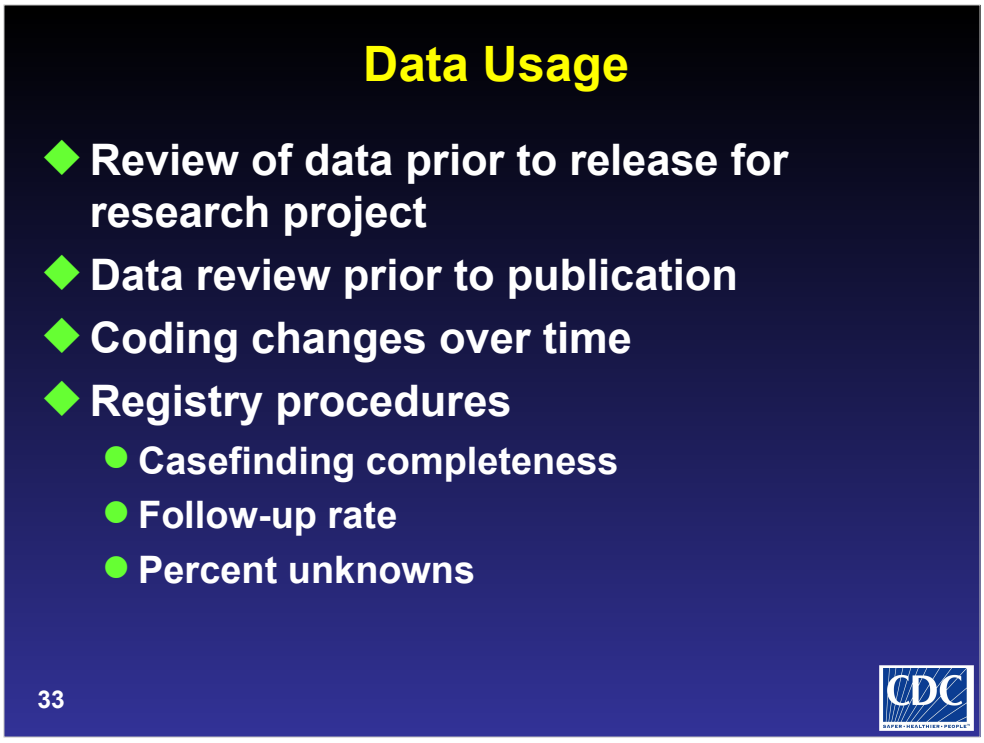
- Breast, lung, colorectal, corpus uteri, cervix, and/or thyroid for Hispanic women.
- Breast, lung, colorectal, corpus uteri, lymphoma, and/or pancreas for African American women.
- Prostate, colorectal, lung, lymphoma, kidney/renal pelvis, and/or bladder for Hispanic men.
- Prostate, colorectal, lung, lymphoma, kidney, and/or oral cavity for African American men.



NPCR, SEER, and NAACCR have set standards and criteria for the maximum percentage of cancers of unknown origin and death certificate only. These percentages affect the overall frequency distribution of the state's annual incidence rates.

In addition to these general patterns, site-specific data patterns need monitoring. It is extremely important that the central registry obtain accurate census information on the population within its geographic area, because the race distribution of the population will affect both the incidence counts and the incidence rates for the geographic area in general. Understanding the race and ethnicity distributions in the population will help make sure that there are no anomalies in cancer registry aggregate data.

Similarly, understanding the age distribution within the state's population will help account for any variability in cancer incidence when the state's data is compared to national data or to other states.



Finally, independent eyes looking at the data—those of researchers *using* the data—can find inconsistencies and issues in both individual and aggregate data that may not be seen when individual abstracts are processed through quality control methods. Obviously, if the researchers identify significant issues, the overall quality of the data becomes a cause for concern, but the feedback from researchers on incongruities in the data (in contrast to true errors) can only strengthen the application of computer edits and visual editing procedures.

When the data are requested by a researcher, it is a good idea for the registry to review the data for discrepancies prior to releasing it. At this point, corrections in the data only make it better for research.

When the research is about to be published, the registry staff should request to review the data once more, to assure that the researchers have not come to invalid conclusions based on data problems.

In addition to the data itself, the registry should provide any researcher with information pertinent to the data. This includes—

- Coding changes over time—there may have been some stage migration due to advances in medical diagnostics, new morphology codes, or different ways of coding treatment over the period for which the data were requested.
- Estimates of casefinding completeness. As previously noted, incomplete casefinding can skew the findings in some primary sites.
- If the registry includes follow-up data, information on the successful follow-up rates by patient age should be provided to the researchers, particularly if they are analyzing survival trends or other outcomes measures.
- Percent unknowns by data field—here again, a large percentage of unknown values in a critical data field can confound the researcher's findings. The researcher and the registry must work together to determine ways to counteract a high percentage of unknowns.

(Go to Module 3, Part I Section D.)

The findings and conclusions in this presentation are those of the authors and do not necessarily represent the views of the Centers for Disease Control and Prevention.

For information about CDC's
Cancer Prevention and Control Programs
and the
National Program of Cancer Registries

Please visit www.cdc.gov/cancer/npcr

35